587 Background

[Most of this should be familiar ... ]

(AS= Alon-Spencer, JLR= Janson-Luczak-Ruciński, RG= Bollobás’ Random Graphs, MR= Molloy-Reed.)

Notation

\(N = \{0, 1, \ldots \}\), \(P = \{1, 2, \ldots \}\), \([n] = \{1, \ldots , n\}\)

\((n)_k = n(n-1) \cdots (n-k+1) \ (= k!(\frac{n}{k}) )\)

\(2^X = \{ \text{subsets of } X \}\)

\(\binom{X}{k} = \{ k\text{-subsets of } X \}\)

For \(f, g\) functions (of whatever):

\(f = \mathcal{O}(g)\) means \(|f| \leq C|g|\) with \(C\) a constant;

\(f = \Omega(g)\) means \(|f| \geq C|g|\) with \(C\) a positive constant;

\(f = \Theta(g)\) (also written \(f \asymp g\)) means \(f = \mathcal{O}(g)\) and \(f = \Omega(g)\);

\(f = o(g)\) (or \(f \ll g\)) means \(f/g \to 0\), \(f = \omega(g)\) (or \(f \gg g\)) means \(|f/g| \to \infty\),

and \(f \sim g\) means \(f/g \to 1\), all as something tends to something or other;

and \(f \approx g\) means nothing precise.

Warning: we’ll often (as above) use “hidden parameters,” i.e. underlying parameters that we suppress in our notation. This may be slightly confusing initially, but will eventually make our lives considerably easier.

Probability

I won’t try to do this for real here, but a few items of particular interest for us are mentioned below. See Ross’s Introduction to Probability Models for a nice review of basics (if you need them). Beyond these, Ross and Feller are good, clear introductions to probability. (Durrett is also nice, but significantly harder.) Actually, the first chapters of \(\text{RG}\) and \(\text{JLR}\) give enough background for most of what we’ll do, and there’s also a little in MR.

Hypergraphs

For us a hypergraph is simply a collection \(\mathcal{H}\) (possibly with repeats) of subsets (called edges) of some finite (vertex) set \(V\). (Note this notation differs slightly from that in AS, p.7.) We say \(\mathcal{H}\) is \(r\)-uniform if each of its edges has size \(r\); \(d\)-regular if each of its vertices has degree \(d\) (i.e. belongs to exactly \(d\) edges); and simple (or nearly-disjoint) if \(|A \cap B| \leq 1\) whenever \(A, B\) are distinct edges.
Random variables (r.v.’s)

Most of our r.v.’s $X$ are discrete (that is, they take values in some finite or countable set), and are therefore specified by the probabilities $P(X = x)$. Some authors require r.v.’s to be real-valued, but we don’t; e.g. we might see r.v.’s whose values are vectors or graphs. As is usual, we use $\mu_X$ or $\mathbb{E}X$ for the mean and $\sigma_X^2$ or $\text{Var}(X)$ for the variance of a real-valued r.v. $X$.

Distributions we’ll tend to see: 1

$X \sim \text{Ber}(p)$ means $X$ is Bernoulli (or a Bernoulli r.v.) with parameter $p$; that is, $P(X = 1) = p$, $P(X = 0) = 1 - p$. If $A$ is an event then its indicator is the Bernoulli r.v. $1_A$ given by

$$1_A = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise.} \end{cases}$$

$X \sim \text{Bin}(n,p)$ means $X$ is binomial with parameters $n,p$; that is, $X = \sum_{i=1}^n X_i$, where the $X_i$’s are independent Ber($p$) r.v.s. Thus $X$ is the number of successes in $n$ independent trials, each with probability of success $p$, and

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \quad 0 \leq k \leq n.$$ 

$X \sim \text{Geom}(p)$ means $X$ is geometric with parameter $p$; that is, $P(X = i) = (1 - p)^{i-1}p$. So we may regard $X$ as the number of trials until success if the trials are independent and each succeeds with probability $p$.

$X \sim \text{Po}(\mu)$ means $X$ is Poisson with mean $\mu$; that is,

$$P(X = k) = e^{-\mu} \mu^k / k! \quad \text{for } k \in \mathbb{N}.$$ 

[Little exercises: (a) find the means and variances of $X \sim \text{Bin}(n,p)$ and $Y \sim \text{Geom}(p)$, and the variance of $Z \sim \text{Po}(\mu)$; (b) what can you say about $X + Y$ if $X$ and $Y$ are independent with $X \sim \text{Po}(\mu)$ and $Y \sim \text{Po}(\lambda)$?]

$X \sim \text{N}(0,1)$ means $X$ is standard normal, i.e. has density function

$$\phi(t) = \frac{1}{\sqrt{2\pi}} e^{-t^2/2}$$

and (therefore) distribution function

$$P(X < x) = \Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-t^2/2} dt.$$

Then $\mu_X = 0$, $\sigma_X^2 = 1$, and one can show that as $x \to \infty$,

$$(P(X \geq x) = \quad 1 - \Phi(x) \sim x^{-1} \phi(x) = \frac{1}{x \sqrt{2\pi}} e^{-x^2/2}).$$
More generally, \( X \sim N(\mu, \sigma) \) has density function \( \frac{1}{\sqrt{2\pi}} e^{-(t-\mu)^2/(2\sigma^2)} \) and (consequently) mean \( \mu \) and variance \( \sigma^2 \).

And we might also see: \( X \sim \text{Exp}(\lambda) \) means \( X \) is exponential with parameter \( \lambda \) \((0 \leq \lambda < \infty)\); that is,

\[
\mathbb{P}(X > x) = e^{-\lambda x} \quad x \geq 0.
\]

**Estimates**

*Stirling’s formula:*

\[
n! \sim \sqrt{2\pi n} \left(\frac{n}{e}\right)^n.
\]

(Even better estimates exist, but we’re unlikely to need them; more often than not we don’t even care about the \( \sqrt{2\pi n} \).)

**Approximation by powers of \( e \)**

For any \( x \in \mathbb{R} \) (often \( x < 0 \)), \( 1 + x < e^x \). This doesn’t give away much if \( x \) is small, and in such cases replacing \( 1 + x \) by the easier-to-work-with \( e^x \) should be more or less a reflex. Sometimes we also need a reverse inequality (e.g. to test the accuracy of \((1 + x)^n \approx e^{xn}\)); for this you can use something like \( 1 + x > e^x - x^2 \), valid at least for \(|x|\) slightly small (exercise to work out more precisely where it’s okay).

**Binomial coefficients**

We use these constantly, so you should get comfortable with them; in particular it might help to go through the easy verifications of the following statements.

For any \( n, k \),

\[
\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \frac{n^k}{k!} \leq \left(\frac{en}{k}\right)^k.
\]

If \( k = o(n^{2/3}) \), then

\[
\binom{n}{k} \sim \frac{n^k}{k!} e^{-k^2/(2n)}
\]

and in particular

\[
\binom{n}{k} \sim \frac{2^k}{k!}
\]

if \( k = o(\sqrt{n}) \). If \( k = \alpha n \) with \( \alpha > 0 \) fixed, then \( \binom{n}{k} \approx 2^{H(\alpha)n} \), where \( H = H_2 \) is binary entropy: \( H(\alpha) = -\alpha \log \alpha - (1 - \alpha) \log(1 - \alpha) \) (where \( \log = \log_2 \)).

More precisely (and usually more precisely than we need),

\[
\binom{n}{k} \sim [2\pi\alpha(1 - \alpha)n]^{-1/2} 2^{H(\alpha)n}.
\]
If $k$ is only assumed to be close to $\alpha n$, then we get approximations of this estimate; for instance, $k \sim \alpha n$ gives $\log \binom{n}{k} \sim H(\alpha)n$, but not (1). (Note “$\log f \sim \log g$” is much weaker than “$f \sim g$”.)

Large deviations

For independent r.v.’s these are treated in great detail in Appendix A of AS. Chapter 2 of JLR is also good (less detailed but more general). We mention just a prototypical example: for $X \sim \text{Bin}(n, \frac{1}{2})$ and $k > 0$,

$$P(X \geq n/2 + k) = 2^{-n} \sum \binom{n}{i} : \binom{i}{n} > n/2 + k < e^{-2k^2/n}.$$  

(Tiny exercise: this is equivalent to AS, Theorem A.1.)

Inclusion-exclusion

For events $A_1, \ldots, A_n$, $A_I := \cap_{i \in I} A_i$ and $S_k := \sum \{P(A_I) : I \in \binom{[n]}{k}\}$,

$$P(\cap_{i \in \mathcal{I}} A_i) = \sum \{(-1)^{|I|}P(A_I) : \mathcal{I} \subseteq [n]\} = \sum_{k=0}^{n} (-1)^k S_k$$

and, more generally (K. Jordán’s formula),

$$P(\text{exactly } r \text{ of the } A_i \text{ occur}) = \sum_{k=r}^{n} (-1)^{k+r} \binom{k}{r} S_k.$$

Moreover the partial sums here alternately over- and underestimate the left hand side; succinctly:

$$(-1)^{l-r} \sum_{k=r}^{l} (-1)^{k+r} \binom{k}{r} S_k \geq 0$$

for each $l \geq r$. When $r = 0$ (corresponding to I-E), these are the Bonferroni Inequalities; in general I assume they’re due to Jordán.

Convergence

A sequence $\{X_n\}$ of $\mathbb{R}$-valued r.v.’s converges in distribution to the r.v. $X$, written $X_n \xrightarrow{d} X$, if $P(X_n \leq x) \rightarrow P(X \leq x) =: F(x)$ for each point of continuity $x$ of $F$. (We also write, e.g., $X_n \xrightarrow{d} N(0,1)$ when $X_n \xrightarrow{d} \text{Bin}(n, \frac{1}{2}) \sim N(0,1)$.) Don’t worry too much about the definition: we will almost always (or always) have either

(i) $X$ is normal (so every $x$ is a point of continuity); or

(ii) $X$ and the $X_n$’s are discrete (probably integer-valued, in which case $X_n \xrightarrow{d} X$ just means that $P(X_n = x) \rightarrow P(X = x)$ for each fixed $x$.}
E.g. (exercise): if $X_n \sim \text{Bin}(n, c/n)$ with $c$ fixed, then $X_n \xrightarrow{d} \text{Po}(c)$.

Similarly (not an exercise): if $X \sim \text{Bin}(n, p)$ with $np(1-p) \to \infty$, then 

$$(X - np)/\sqrt{np(1-p)} \xrightarrow{d} N(0, 1).$$

(Note the hidden parameter: $p$ is $p(n)$.)

This is a special case of both the DeMoivre-Laplace Theorem (see e.g. RG, p.13 for a very complete statement, the last couple lines of which are what we want) and the Central Limit Theorem.